Weakly Supervised rPPG Estimation for Respiratory Rate Estimation

Jingda Du, Si-Qi Liu, Bochao Zhang, Pong C. Yuen
Department of Computer Science, Hong Kong Baptist University, Hong Kong
csjddu, siqiliu, csbczhang, pcyuen@comp.hkbu.edu.hk

Abstract

Recent studies demonstrate that respiratory rate can be estimated from skin videos through analyzing the frequency domain attributes of their remote photoplethysmography (rPPG). However, respiration is not always periodic so the frequency attributes of rPPG may not accurately estimate the respiratory rate. In this paper, we proposed an end-to-end network to estimate both rPPG signals and respiratory rates from facial videos. Since only breathing waves are available in the Remote Physiological Signal Sensing track2 competition, to preserve the respiratory pattern in rPPG estimation, rPPG signals pre-estimated by chrominace-based methods and modulated by breathing waves are used as weak labels for supervision. To adapt to the large differences between training and testing data, in terms of recording environment and subjects behavior, we also involved customized adversarial training on feature extractor to minimize the domain gap. In the competition, our model achieved 7.56 bpm MAE and ranked the second place.

1. Introduction

The monitoring of physiological signals like heart rate, respiratory rate is essential for health condition analysis. Common medical measurements of heart rate and respiratory rate rely on Electrocardiography (ECG) and Photoplethysmograph (PPG). Both ECG and PPG require skin contact during measurement and may cause uncomfortable sensation of testees in long-term monitoring. To overcome this problem, many researchers have been exploring on remote photoplethysmography (rPPG), which can remotely estimate physiological signals through extracting pulse-induced subtle light absorption variation in facial videos [3, 15, 2, 19, 13, 12].

Most of the existing methods do not directly estimate respiratory rate from facial videos but firstly estimate rPPG signals and then analyze the frequency domain attributes of rPPG signals for respiratory rate estimation. Because respiration is related to frequency and amplitude variation of rPPG which is subtle in facial videos, it's hard to estimate respiratory rate accurately from facial videos. Accurate remote respiration rate estimation requires to extract respiration patterns from rPPG signals. As shown in the Figure 1.

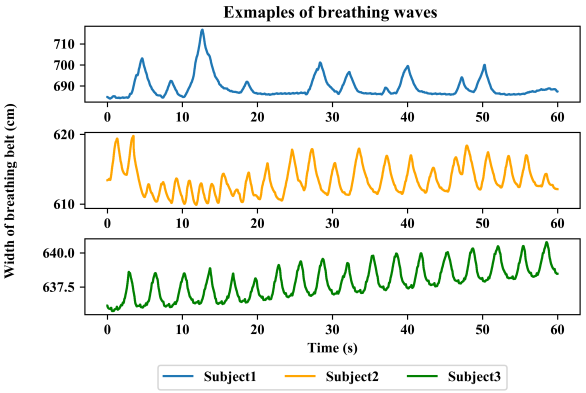

Figure 1. Different types of breathing waves. The breathing waves are provided in the competition as ground truth.

In fact, respiration can influence heart rate and blood volume change, which correspondingly modulate the frequency and amplitude of rPPG signals. Both frequency modulation (FM) and amplitude modulation (AM) in rPPG signals can be used to estimate respiratory rate [4]. The extraction of the two modulations requires high quality rPPG signals. For deep learning based approach, ground truth PPG signals are needed for training. While in the competition, only breathing waves are available so it is hard to estimate the respiration information directly from facial videos. Besides, the training dataset and test dataset have large difference on participants' behaviours and recording environments. Different from the well-controlled training dataset, the participants' motion and lighting conditions in the test dataset is more complex.

To address these problems, we proposed an end-to-end domain adaptive network and an weakly supervised learning strategy by using the breathing waves modulated rPPG signals for respiratory rate estimation. Our contributions include:

- We proposed an end-to-end network for respiratory rate estimation via analyzing respiration patterns in rPPG signals.
- We modulated pre-estimated rPPG signals with breathing waves as weak labels to obtain pseudo rPPG signals with respiration pattern for rPPG estimator training. Besides, mask negative Pearson correlation loss is proposed to enhance the respiration pattern learning for rPPG estimator.
- An adversarial learning strategy is designed to reduce the domain gap in terms of recording environment and motion interference.

2. Related Work

Remote Photoplethysmography Estimation. The remote photoplethysmography (rPPG) can estimate pulse-induced blood volume change through extracting related light absorption variations. The rPPG signals contain multiple physiological signals like heart rate, respiratory rate, heart rate variability. To estimate rPPG signals from facial videos, Blind Source Separation (BSS) techniques are applied to separate rPPG signals from noise [9, 14]. These methods can extract periodic rPPG signals and effectively remove the noise in RGB channels. However, subjects' motion induced noise, especially periodic noise can disturb BSS-based methods' estimation. To estimate rPPG signals under subjects' motion, a chrominace-based method was proposed to estimate rPPG through combining RGB channels under white lighting assumption and utilizing skin standardization to keep the white lighting [3]. The skin standardization requires precognition on the skin-tone information and skin-tone mismatch may cause inaccurate rPPG estimation. [17] proposed to estimate rPPG signals by combining the normalized RGB channels.

Recently, deep learning based methods have been proposed to learn accurate and robust mapping from facial videos to rPPG signals. [15] proposed an end-to-end, two-stage 2D CNN model to estimate rPPG signals and average heart rates. This method outperforms traditional methods but also can be affected by video compression. To estimate rPPG signals accurately, spatial temporal convolution based methods are proposed to utilize temporal information contained in facial videos to estimate rPPG signals [19, 12]. Besides, [13] proposed the multi-scale spatial-temporal map to represent the raw facial video and the cross-verified scheme to adaptively split noise from physiological features.

rPPG-based Respiratory Rate Estimation. The high quality rPPG signals contain various physiological features and rPPG-based respiratory rate estimation is a very prospective application. [10] proposed to extract inter-beat-interval (IBI) signals from rPPG signals and analyze IBI signals' frequency attributes for respiratory rate estimation. But IBI-based respiratory rate estimation demands more strict constraints. The peak positions of rPPG signals should be accurate. The respiration should be periodic. [2] estimates rPPG signals from facial videos and analyzes peak values in the power spectrum density (PSD) of rPPG signals to estimate heart rate and respiratory rate. The two methods both adopted frequency analysis as the foundation of respiratory rate estimation, however, for situations with non-periodic respiration, errors may occur.

Domain Adaptation. Supervised learning can utilize large-scale data to train powerful models and achieve high performance on the independent and identical distributed test dataset. However, when large domain shift exists between source domain and target domain, only using the source domain for training may lead to a biased model and the performance on target domain will severely decrease. Learning domain-invariant feature representation can effectively reduce the influence of domain shift and decrease the performance drop on test dataset [20]. [16] proposed the correlation alignment to minimize the divergence between source domain and target domain. [6] proposed to train the encoder to learn domain-invariant features by utilizing the features for both classification and reconstruct samples to target domain. [5] utilized adversarial training strategy between feature extractor and domain classifier which allows feature extractor to learn domain-invariant features from source domain and target domain. Normalization layers like batch normalization [7], group normalization [18] are widely applied in neural network architecture and their mismatch in the test time may induce lower performance. [11] assumes that the normalization layers learn domain information and neural network layers learn task-related information. Under this assumption, the AdaBN was proposed to adjust normalization layers' parameters according to both the source domain and target domain to overcome the influence of domain shift. Different from usual domain gap in image classification like corruption types, picture styles and so on, domain shift in our case combines lighting variation, motion variation and the shift of main range of heart rate and respiratory rate. It increases our difficulty when doing domain adaptation since the diminishing domain gap may cause the diminishing physiological information. In this paper we customize the domain adaptation process and reduce the domain gap of physiological signals between training dataset and testing dataset, thus reduce the influence of domain adaptation to our main objective.

3. Methodology

In this section, we firstly introduce how the proposed method estimates rPPG signals and respiratory rate from facial videos, then unveil the pseudo rPPG labels generation process using breathing wave, and finally depict how to adapt the complex domain shift in testing data. Figure 2

shows the whole framework of the proposed method. Our model adopts both training data and test data as input to learn domain-variant physiological features for respiratory rate estimation. And we utilize pre-estimated rPPG signals, breathing waves and domain information to generate ground truth for model training.

3.1. An End-to-End Network for Respiratory Rate Estimation

The rPPG can reflect blood volume change by analyzing subtle light absorption variation in facial videos, while the blood volume can be influenced by several physiological behaviours. According to previous rPPG-related papers [10, 2], the information of both heart rate and respiratory rate is contained in the rPPG signal. Because the rPPG is subtle in the facial video and the respiration information is weak or even fuzzy in rPPG, it is hard to directly estimate respiratory rate from the facial video by a DNN model. Therefore, we split the respiratory rate estimation into three steps: (1) extracting features that contain accurate and domain-invariant physiological information, (2) estimating rPPG signals with respiration pattern from facial videos, (3) estimating respiratory rate from estimated rPPG signals. And the three steps correspond to feature extractor, rPPG estimator and domain classifier in the model. Before introducing the model, we first introduce model's input which is also important for respiratory rate estimation

In the previous rPPG estimation methods, both raw facial videos and corresponding multi-scale spatial-temporal map (MSTmap) [13] are used as input, while in this competition, to protect participants' privacy, the facial videos provided in the training dataset contains dynamic mosaics on eyes and mouth. Besides, 68 landmarks are provided as auxiliary information. The mosaics' irregular variation and movement will disturb rPPG estimation if raw videos are selected as input. To preserve physiological information without mosaics' disturbance, we selected 3 regions of interest in each facial video and generated a MSTmap with mean values of RGB channels as shown in the Figure 3. We also smoothed the facial landmarks to reduce the noise and disturbance caused during their detection . In the test stage, the MSTmap is generated with the same RoIs' selection and the facial landmarks are detected by the open source package face-alignment[1].

To estimate respiratory rate from the MSTmap, we utilized 2D convolution, 2D transposed convolution, fully connected layer and batch normalization layer to build our model and its architecture is shown in Figure 4. All activation functions in the model are ReLU [1].

Firstly we used the convolution layers to extract the physiological features and down-sampling operation is applied to reduce noise in extraction. To reduce the information loss and keep accurate physiological features, the convolution with stride equal to 2 is selected for down-sampling. Different from heart rate estimation, respiration is related to the modulations of rPPG. Accurate rPPG peaks' location or amplitude variation is required for respiratory rate estimation. To obtain rPPG signals with more details, we used the 2D transposed convolution to up-sample the physiological features. And a 1×1 convolution is applied to convert physiological features into the rPPG signal. The labels and loss functions for rPPG estimation supervision will be introduced in the next sub-section. The respiratory rate estimator is composed of three fully connected layers and can globally extract respiration pattern for estimation. The ground truth respiratory rate is equal to the number of peaks in the breathing wave. We selected the mean square error (MSE) as the loss function

$$L_{RR} = \frac{\sum_{i=1}^{T} (RR^i_{pred} - RR^i_{gt})}{N} \quad (1)$$

RR_{pred} is respiratory rate estimated by model, RR_{gt} is the ground truth respiratory rate and N is the number of samples in one batch.

3.2. Weak Labels for rPPG Estimation

In the previous deep learning based rPPG estimation methods, the PPG or ECG is provided for supervision. Especially adopting PPG as ground truth and the negative Pearson correlation as loss function, is a common practice in rPPG estimation tasks. However, in this respiratory rate estimation competition, only breathing waves are provided. On one hand, without PPG and ECG deep learning based methods can can easily diverge or learn nothing but memorize, due to high implicity of our task. On the other hand, it is hard for traditional methods to estimate accurate rPPG signals containing respiration pattern in test dataset with strong motion disturbance. While in the training dataset, the facial videos are captured under well-controlled environment with little participants' motion. So traditional methods can also estimate the rPPG signal with limited error and we used [3] to generate pre-estimated rPPG signals from training dataset.

Therefore, we proposed to utilize breathing waves to modulate the pre-estimated rPPG signals as the ground truth to supervise rPPG estimation. Our idea of pseudo rPPG generation is shown in the Figure 5. In the pseudo labels' generation, the pre-estimated rPPG signal is firstly normalized to the same amplitude. Then the breathing wave is normalized and vertical flipped because respiration usually decreases the amplitude of rPPG signals. The pseudo label is generated by multiplying the normalized two signals and presents amplitude variation during respiration. When we train the model to estimate rPPG signals, we firstly used the

[1] https://github.com/1adrianb/face-alignment

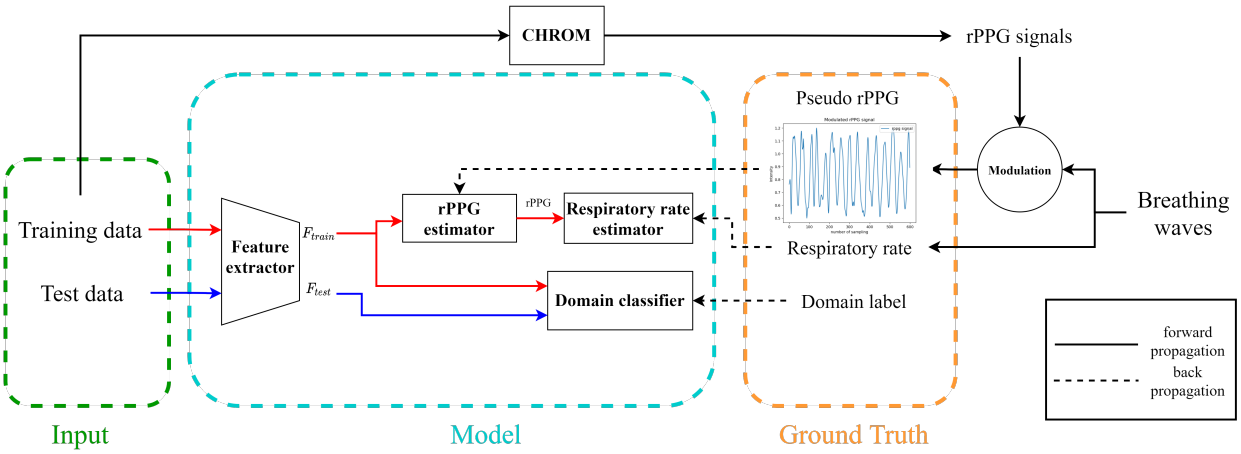

Figure 2. Overview of proposed method. The training data and test data are both provided for training. The feature extractor is trained to obtain physiological features for rPPG signal and respiratory rate estimation. The adversarial training is used between feature extractor and domain classifier for learning domain-invariant physiological features. The rPPG estimator takes F_{train} as input and estimate rPPG signals which contain respiration patterns. The respiratory rates are estimated by the estimator. The pseudo rPPG is generated through using breathing waves to modulate pre-estimated rPPG signals.

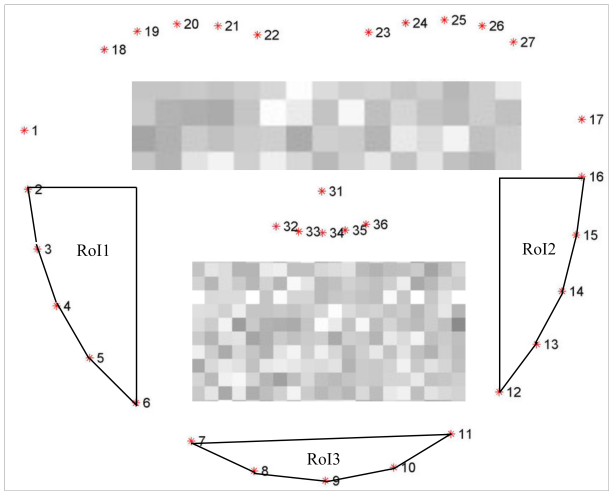

Figure 3. We selected three RoIs for MSTmap generation to reduce the influence of dynamic facial mosaics contained in training dataset.

negative Pearson correlation as a loss function

$$L_{rPPG} = 1 - \frac{Cov(est, gt)}{\sqrt{Cov(est, est)\,Cov(gt, gt)}} \qquad (2)$$

the est presents the rPPG signal estimated model and gt presents the corresponding pseudo label. Cov in the equation is the function. Meanwhile, to increase the amplitude modulation's weight in training, we proposed the mask neg-

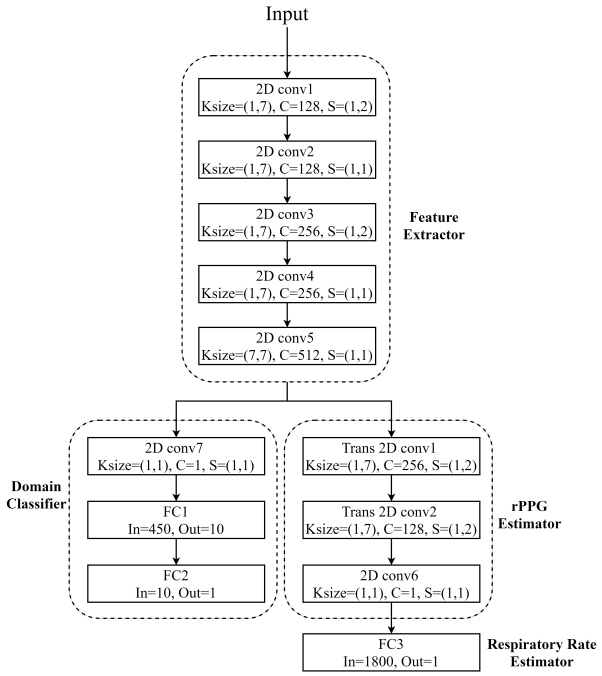

Figure 4. The figure presents the details of our proposed method. The Ksize represents kernel size, C represents output channel number, and S represents the stride of convolution operation.

ative Pearson correlation loss function

$$L_{trend} = 1 - \frac{Cov(est_m, gt_m)}{\sqrt{Cov(est_m, est_m)\,Cov(gt_m, gt_m)}} \qquad (3)$$

the est_m presents the masked estimated rPPG signal and

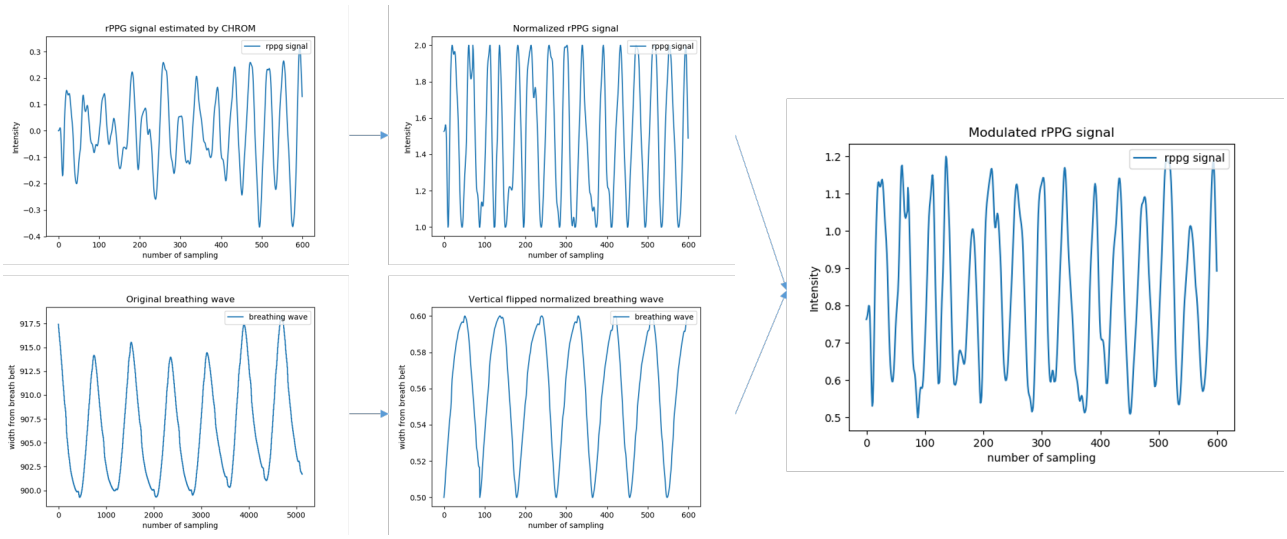

Figure 5. An example of pseudo labels' generation. Both the rPPG signal and the breathing wave are required to normalized into a fixed range. Through multiplying normalized two signals, the generated pseudo rPPG contains respiration-induced amplitude modulation. In the generated pseudo rPPG signal, the respiration-related amplitude variation can be clearly captured.

gt_m presents the masked pseudo label. The binary masks are generated according to the vertical flipped normalized breathing waves and only top 20 percent values are 1 in the masks. The binary masks preserve the correlation between peak regions in rPPG and optimizing the correlation can emphasize respiration-induced amplitude modulation in rPPG. The final loss for rPPG and respiratory rate estimation is

$$L_{phys} = L_{RR} + \alpha L_{rRRG} + \beta L_{trend} \tag{4}$$

where α and β are hyperparameters.

3.3. Domain Adaptation

In this respiratory rate estimation competition, the training dataset and test dataset have large difference on both respiratory rate range and disturbance contained in facial videos. Resampling of training data can effectively reduce the influence of respiratory rate range difference. To overcome the influence of different disturbance, ordinarily, applying adversarial training between the feature extractor and domain classifier can help feature extractor learn domain-invariant features. However, when the main task related features can be classified, the domain classifier may disturb the feature extractor's learning but not help to learn domain-invariant features.

In our case, the respiratory rate distributions in training dataset and test dataset are different. Because, the participants in the training dataset are just staying still while the participants in test dataset are exercising in the gym. To adapt to the higher respiratory rate in exercise, test dataset has larger respiratory rate range than training dataset. The

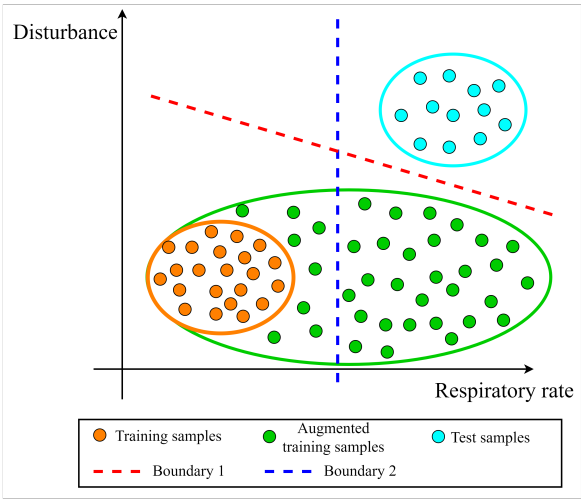

Figure 6. The example of domain distribution of training dataset, test dataset and augmented training dataset. The training dataset and test dataset have different respiratory rate range which can be used for domain classification. The decision boundary of domain classifier on training dataset and test dataset may be similar to boundary 1 and the feature extractor may not adapt to the disturbance difference. While the augmented training dataset cannot be differentiated from test dataset by respiratory rate range and the disturbance-induced domain shift can be reduced in the physiological feature space.

example of domain distribution is shown as the Figure 6. Therefore, we up-sampled the MSTmaps in the training data with both 1.5 and 2 times for frequency augmentation.

And this kind of augmentation can reduce the respiratory rate and heart rate difference between the two datasets. After the frequency augmentation, the domain classifier forces the feature extractor to learn a domain invariant feature that can adapt the difference between training and testing dataset. The domain classifier is trained by the Binary Cross Entropy (BCE) function

$$L_{dc} = y \log x + (1-y) \log(1-x) \tag{5}$$

x presents the predicted domain class and y presents the domain label.

4. Experiments

In this section, we experimented our method on the track2 competition and some ablation studies are provided to validate the effectiveness of our method.

4.1. Datasets

In the track2 competition, a large scale training dataset and a challenging test dataset are provided for remote respiratory rate estimation. The training dataset contains 995 one-minute facial videos captured from 100 participants and all the videos are recorded with 1920 $\times$ 1080 resolution at 30 FPS. To protect the privacy of participants, all facial videos in the training dataset are covered by two mosaics on the eyes' region and mouth region as shown in the Figure 3. And all facial videos in the training dataset have 68 facial landmarks' coordinates provided. Each video has a corresponding breathing wave recorded at 256Hz provided as the ground truth and the ground truth respiratory rate can be gained by counting the peaks in the breathing wave.

In our experiments, we divided the training dataset into a subtraining dataset and a validation dataset. The subtraining dataset contains the first 90 participants' facial videos and the validation dataset contains the last 10 participants' facial videos.

The test dataset contains 283 one-minute facial videos captured from 10 participants and all the videos are recorded with 1280 $\times$ 720 resolution at 30 FPS. The videos are captured in the gym under running and cycling exercising modes and contain disturbances from subjects' motion and complicated ambient lighting.

4.2. Implementation Details

Pseudo rPPG label generation. As mentioned in section3, the pseudo rPPG labels are generated by modulating pre-estimated rPPG signals with the corresponding breathing wave. The pre-estimated rPPG signals are estimated by CHROM [3] with the mean RGB values of RoI3. The landmarks for mean RGB values' calculation are all smoothed by the mean filter. The changing range of peak amplitude in the pseudo rPPG labels is 1.0-1.2. The breathing waves are down-sampled to 30Hz.

Data augmentation. To reduce the difference between training dataset and test dataset, we firstly generated 1.5 times and 2 times up-sampled MSTmaps by resampling the MSTmaps extracted from two adjacent facial videos. We also flipped all the training samples in the time dimension to increase the diversity while training.

Domain adaptation. To reduce the domain gap induced performance drop on target domain, we utilized adversarial training between feature extractor and domain classifier. The domain labels are generated according to belonging dataset. As an auxiliary task for respiratory rate estimation, the domain adaptation training has smaller learning rate than the main task.

Training setting. In the training stage, the model is implemented with PyTorch and trained on Nvidia V100. We used Adam optimizer [8] with weight-decay equals 1e-4. The learning rate for physiological model is 1e-4 and the learning rate for domain adaptation model is 1e-6. The model is trained on the subtraining dataset for 20 epochs and the model weight with lowest validation error is selected for test submission. To keep the numerical stability of loss functions based on Pearson correlation in the training stage, we added a small number 1e-8 to the denominator to prevent NaN loss values. In the loss function L_{phys}, we set α = 0.1 and β = 2.

4.3. Results and Analysis

We trained our proposed method with different settings and the results are shown in the Table 1. In settings column, 1x, 1.5x and 2.x represent the resampling times of MSTmap. The mean absolute error (MAE), root mean square error (RMSE), and Pearson correlation coefficient (R) are provided.

From the results, we can find that upsampling can help learning information from high respiratory rate samples and improve the performance on the test dataset. In fact, the difference of respiratory ranges between training dataset and test dataset can influence model's performance to a large extent. In the physiological signal' estimation, resampling the training dataset covers the whole reasonable range, which can help improve performance on unseen application scenarios.

Besides, the test results also show that the domain adaptation under data augmentation can help the model adapt to the test samples and estimate respiratory rate robustly. We only use the domain adaptation strategy on augmented training dataset. Because when training dataset and test dataset respiratory ranges have little or no interaction, domain classifier may intend to utilize respiratory rate as domain classification criterion and feature extractor may not adapt to the disturbance in test dataset.

Table 1. The test results of our proposed method under different settings.

Settings	MAE (bpm)	RMSE (bpm)	R
1x	11.84	14.45	0.07
1x + 2x	8.70	10.86	-0.003
1x + 2x + DA	8.00	9.73	0.15
1x + 1.5x + 2x + DA	7.85	9.56	0.14
Fusion	7.55	9.04	0.01

Finally, we merged the results from two models with the domain adaptation strategy by calculating the mean of two test results and achieved our best result.

5. Conclusion

In this paper, we proposed an end-to-end convolution neural network for respiratory rate and rPPG estimation. To estimate rPPG signals with respiratory pattern, we modulated pseudo rPPG with pre-estimated rPPG signals and breathing waves. We also utilized domain adaptation mechanism to reduce the influence of domain shift between training dataset and test dataset when the information related to the main task is indistinguishable after data augmentation. Only using breathing waves may be not enough for high rPPG signal and accurate respiratory rate estimation. Estimating accurate respiratory rate under less constrained conditions using large scale dataset with more physiological signals provided may be one future research direction.

References

[1] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[2] Weixuan Chen and Daniel McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 349–365, 2018.

[3] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[4] Marc-André Fiedler, Micha Rapczyński, and Ayoub Al-Hamadi. Fusion-based approach for respiratory rate recognition from facial video images. *IEEE Access*, 8:130036–130047, 2020.

[5] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

[6] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *European conference on computer vision*, pages 597–613. Springer, 2016.

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[8] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity. In *2011 federated conference on computer science and information systems (FedCSIS)*, pages 405–410. IEEE, 2011.

[10] Xiaobai Li, Iman Alikhani, Jingang Shi, Tapio Seppanen, Juhani Junttila, Kirsi Majamaa-Voltti, Mikko Tulppo, and Guoying Zhao. The obf database: A large face video database for remote physiological signal measurement and atrial fibrillation detection. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 242–249. IEEE, 2018.

[11] Yanghao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu. Adaptive batch normalization for practical domain adaptation. *Pattern Recognition*, 80:109–117, 2018.

[12] Si-Qi Liu and Pong C Yuen. A general remote photoplethysmography estimator with spatiotemporal convolutional network. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 481–488. IEEE, 2020.

[13] Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, and Guoying Zhao. Video-based remote physiological measurement via cross-verified feature disentangling. In *European Conference on Computer Vision*, pages 295–310. Springer, 2020.

[14] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.

[15] Radim Špetlík, Vojtech Franc, and Jirí Matas. Visual heart rate estimation with convolutional neural network. In *Proceedings of the british machine vision conference, Newcastle, UK*, pages 3–6, 2018.

[16] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016.

[17] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard De Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2016.

[18] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[19] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 151–160, 2019.

[20] Han Zhao, Remi Tachet Des Combes, Kun Zhang, and Geoffrey Gordon. On learning invariant representations for domain adaptation. In *International Conference on Machine Learning*, pages 7523–7532. PMLR, 2019.